



**The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:**

**Document Title:** NIJ Recidivism Challenge, 2021, Team Early Stopping, Years 2 and 3

**Author(s):** Team Early Stopping

**Document Number:** 305139

**Date Received:** August 2022

**Award Number:** NIJ Recidivism Forecasting Challenge Winning Paper

**This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.**

**Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.**

NIJ Recidivism Challenge, 2021  
Team Early Stopping, Years 2 and 3

## Variables

We did not engineer many new features, but did convert features to ordered numeric values whenever possible.

Also, before fitting the final model we fit a series of preliminary boosted tree models to impute missing data and to predict secondary outcomes.

For example, to predict Year 2, we first predicted Year 3 and Any Year, and then used out-of-fold predictions for them as input features. The two imputed supplementary targets ended up being the most important variables by a wide margin, which makes sense given their interrelationships with the primary target.

## Models

We used ensembles of gradient boosted tree models from three packages: XGBoost, LightGBM, and Catboost. We used five-fold cross-validation to create out-of-fold predictions, and then a linear model as the second-level model to determine final blending weights. We used Brier Score and Binary Cross Entropy as loss functions and made no special adjustments for Male/Female or for thresholding.

## Future Considerations

The way the challenge was constructed, it is not clear the Year 2 and Year 3 results will be very helpful in practice, since they are conditional on previous years. Is the justice system really that interested in the chance a criminal will recidivate in year 2 or year 3 given they have not recidivated in previous years?

Rather than a binary target for each year, a perhaps more meaningful approach would be to forecast time to recidivism in days or weeks, as in a survival or time-to-event analysis. This would provide much more fine-grained and personal predictions. For example, there is a big difference between someone who has high probability of recidivating in 7-14 days versus 200-300 days, rather than just grouping them all together as "Recidivated in Year 1".

Veronica Comments:

Overall it's vague.

I would go back to the Early Stopping team and ask them to fix their write-up so that we could replicate it?

(e.g. specific modeling decisions and parameters, what software they used, how they used boosted tree models to impute missing data, what changes they made to their model between years 2 and 3 [if any]).